BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 10: e183602
doi: 10.3897/biss.10.183602

OPEN ACCESS

Conference Abstract

Determining the Scale of Herbarium Sheet Images via Ruler Detection

Kenzo Milleville‡, Dieter De Witte‡, Maarten Trekels§, Quentin Groom§, Steven Verstockt‡

‡ IDLab, Ghent University, Ghent, Belgium
§ Meise Botanic Garden, Meise, Belgium

Corresponding author: Kenzo Milleville (kenzo.milleville@ugent.be)

Received: 26 Dec 2025 | Published: 23 Mar 2026

Citation: Milleville K, De Witte D, Trekels M, Groom Q, Verstockt S (2026) Determining the Scale of Herbarium Sheet Images via Ruler Detection. Biodiversity Information Science and Standards 10: e183602. https://doi.org/10.3897/biss.10.183602

Abstract

The digitization and online availability of herbarium specimens have increased the need for automated tools to extract specimen traits at scale. Quantitative trait data, such as leaf area or petal length, are critical for ecology, evolution, and climate change research. In recent years, automated methods such as LeafMachine2 (Weaver and Smith 2023) and Mothra (Wilson et al. 2022) have been developed to measure these traits from images, supporting large-scale studies. However, such image measurements need to be converted from pixels to standardized units (e.g., centimeters) to be useful. Existing methods estimate this conversion factor (CF) by detecting the ruler type and the tick mark locations directly. From these image coordinates, the average pixel distance between them can be calculated to estimate the CF. These methods typically require many annotated samples for each ruler type and tend to perform poorly when applied to unseen or visually distinct rulers.

To overcome these limitations, we present a generic pipeline that accurately predicts the CF with few labeled rulers. The approach leverages the fact that many institutions reuse identical rulers, often differing only in institutional logos. Instead of directly classifying or predicting the tick mark locations, the pipeline uses a similarity-based approach. First, rulers on the page are detected using a YOLOv11 object detection model. Second, these rulers are cropped and normalized, ensuring the longest side is horizontal. After detection, we use Contrastive Language-Image Pre-training (CLIP) to embed each ruler

© Milleville K et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and to retrieve the three most similar ruler candidates from our ground-truth dataset. Next, we use the LightGlue keypoint-matching model to detect corresponding keypoints between the query ruler and each retrieved candidate. This model detects and scores similar keypoint pairs (pixel coordinates) between two input image pairs. The candidate ruler with the highest average score across the 100 best keypoint pairs is selected as the best match. If none of the candidates had at least 10 matching keypoint pairs, we consider the query ruler a “no-match”. Finally, these keypoint pairs are used to estimate a homography, which describes how one image can be transformed to align with another taken from a different viewpoint. This transformation is then applied to the ground truth tick mark locations to calculate the CF for the query ruler.

To evaluate the proposed method, we constructed a new dataset by randomly selecting 400 herbarium sheets from the Global Biodiversity Information Facility (GBIF), PlantCLEF 2020, and Dillen et al. 2019**.** Each ruler was annotated with its bounding polygon and a type. Ruler tick marks were annotated with their coordinate locations and units (e.g., ticks_1cm). The dataset contains a diverse set of 476 rulers from over 40 institutions. We estimated our measurement error at 0.794%, which is in line with the reported 0.8% from LeafMachine2, by calculating the average standard deviation of pixel distances between each tick mark. We evaluated our approach using a leave-one-out cross-validation, treating each ruler crop as a query and the remaining dataset as candidate matches. We also evaluated our method on the 1131 rulers from LeafMachine2, which were annotated with a bounding box around the tick marks (see Fig. 1).

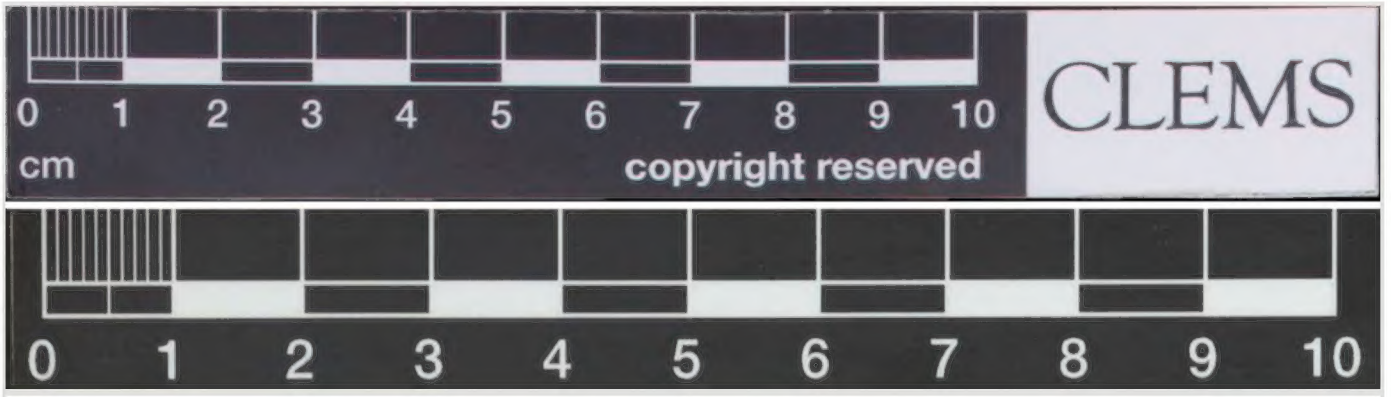

Figure 1.

Top: Sample ruler from our own dataset. Bottom: Sample ruler from the LeafMachine2 dataset. Original images from Clemson University Herbarium, CC BY-NC 3.0.

Rulers without a matching type were excluded for evaluation. The method achieved an average CF error of 0.689% using the best-matching ruler. 94.8% of rulers had a CF error less than 2%. On the LeafMachine2 dataset, performance improved with an average CF error of 0.324%. There, 99.0% of rulers had a CF error less than 2%. We plan to investigate how different annotation strategies affect performance. Table 1 details these evaluation results.

The proposed method offers several advantages over existing approaches. First, it allows an accurate conversion with fewer annotated rulers of the same type, essentially needing only one. Second, the keypoint matching process is robust to partial detections,

occlusions, and low-resolution rulers, as visualized in Fig. 2. Third, only the detection model needs minimal fine-tuning when new ruler types are added to the dataset, as the matching models remain unchanged. Additionally, the prediction can still be accurate when the best-matching ruler is from a different institution (see Fig. 3).

Table 1.

Evaluation results of our pipeline on our dataset and the dataset of LeafMachine2.

Dataset	Amount	Annotation error (%)	Average error (%)	Median error (%)	No match
Own	457	0.794	0.689	0.199	11 (2.40%)
LeafMachine2	1131	0.800	0.324	0.141	46 (4.07%)

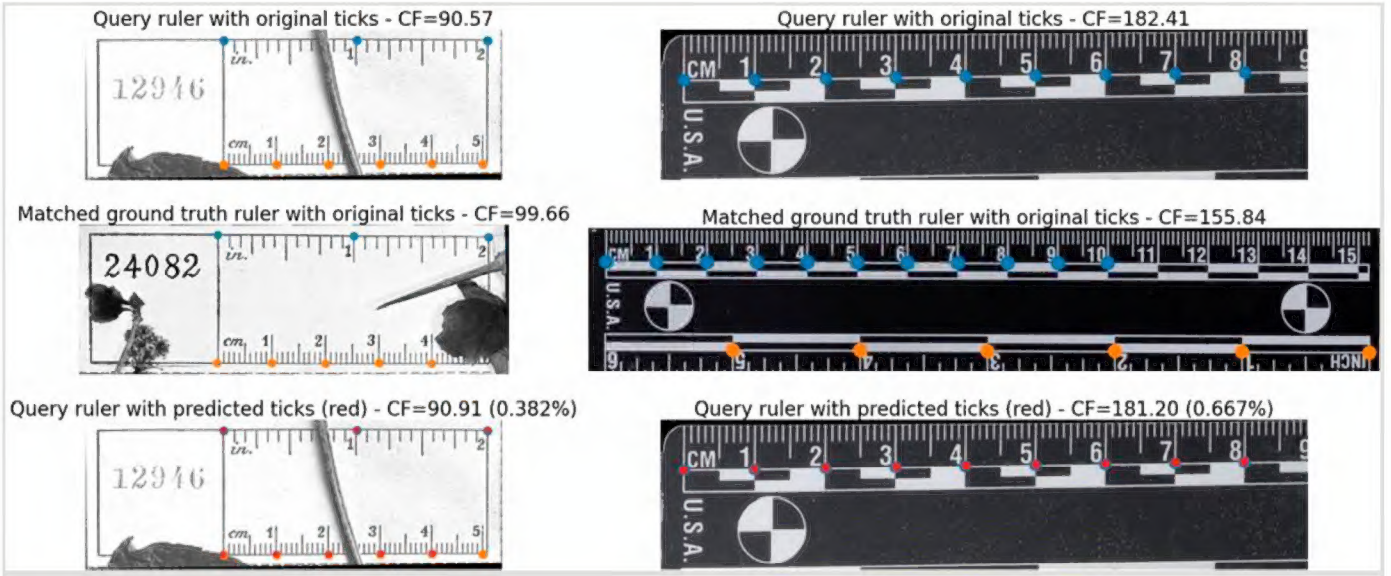
Query ruler with original ticks - CF=90.57
Query ruler with original ticks - CF=182.41
Matched ground truth ruler with original ticks - CF=99.66
Matched ground truth ruler with original ticks - CF=155.84
Query ruler with predicted ticks (red) - CF=90.91 (0.382%)
Query ruler with predicted ticks (red) - CF=181.20 (0.667%)

Figure 2.

Output of our pipeline for two sample rulers. From top to bottom: the query ruler, the best-matching ruler candidate, and the predicted tick marks visualized in red. Original images from the Field Museum of Natural History (CC BY-NC 4.0) and the University of Florida Herbarium (CC BY-NC 4.0).

The main limitation of our method is that it requires at least one annotated ruler of a similar type to result in a good match. It also assumes the tick marks are on a straight line, resulting in poorer performance for curved rulers. Finally, we found that simple rulers, such as straight lines, tend to perform worse (see Fig. 3).

This work represents part of our ongoing research to assist trait extraction at scale. We plan to expand the dataset with additional ruler types, including examples from other natural history collections. We will further refine the method to improve its robustness and generalization across institutions and to explore fallback methods when no match is found. The code and dataset are available on GitHub*[1].

Keywords

computer vision, object detection, trait extraction

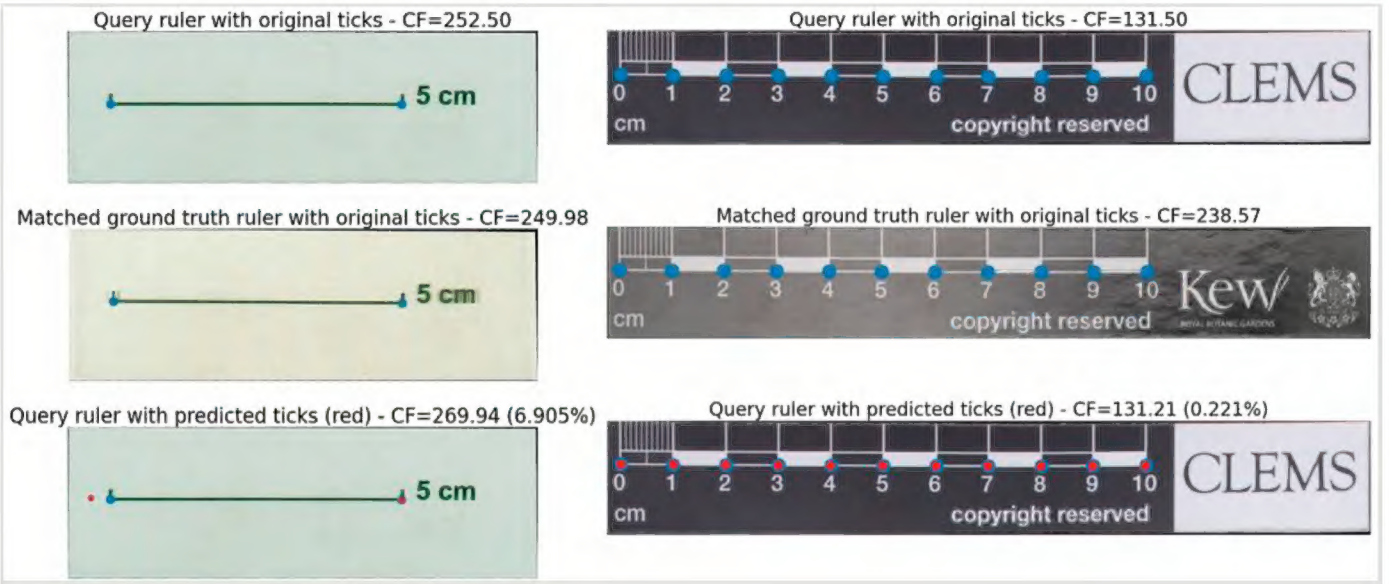
Query ruler with original ticks - CF=252.50
5 cm
Query ruler with original ticks - CF=131.50
0 1 2 3 4 5 6 7 8 9 10
cm
copyright reserved
CLEMS
Matched ground truth ruler with original ticks - CF=249.98
5 cm
Matched ground truth ruler with original ticks - CF=238.57
0 1 2 3 4 5 6 7 8 9 10
cm
copyright reserved
Kew
Query ruler with predicted ticks (red) - CF=269.94 (6.905%)
5 cm
Query ruler with predicted ticks (red) - CF=131.21 (0.221%)
0 1 2 3 4 5 6 7 8 9 10
cm
copyright reserved
CLEMS

Figure 3.

Left: Inaccurate prediction due to a failure in the keypoint matching. Right: Accurate prediction despite the best-matching ruler being from a different institution. Original images from Botanic Garden and Botanical Museum Berlin (CC BY-SA 3.0 DE), Clemson University Herbarium (CC BY-NC 3.0), and Royal Botanic Gardens, Kew (CC BY-4.0)

Presenting author

Kenzo Milleville

Presented at

Living Data 2025

Funding program

This research was supported by the Research Foundation - Flanders (FWO Grant 12A9L25N) and facilitated by the infrastructure provided by DiSSCo Flanders (FWO Project I001625N).

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Dillen M, Groom Q, Chagnoux S, et al. (2019) A benchmark dataset of herbarium specimen images with label data. Biodiversity Data Journal 7 https://doi.org/10.3897/bdj.7.e31817

- Weaver W, Smith S (2023) From leaves to labels: Building modular machine learning networks for rapid herbarium specimen analysis with LeafMachine2. Applications in Plant Sciences 11 (5). https://doi.org/10.1002/aps3.11548
- Wilson R, de Siqueira AF, Brooks S, et al. (2022) Applying computer vision to digitised natural history collections for climate change research: Temperature-size responses in British butterflies. Methods in Ecology and Evolution 14 (2): 372-384. https://doi.org/10.1111/2041-210x.13844

Endnotes

*1 https://github.com/kymillev/herbarium-rulers